This is the version of the article accepted for publication in *Metroeconomica* published Wiley on 7 June 2016, available at: http://dx.doi.org/10.1111/meca.12134
Accepted version downloaded from SOAS Research Online: http://eprints.soas.ac.uk/23157/

A Statistical Equilibrium Approach to the Distribution of Profit Rates*

Ellis Scharfenaker† and Gregor Semieniuk‡

May 18, 2016

Abstract

Motivated by classical political economy we detail a probabilistic, "statistical equilibrium" approach to explaining why even in equilibrium, the equalization of profit rates leads to a non-degenerate distribution. Based on this approach we investigate the empirical content of the profit rate distribution for previously unexamined annual firm level data comprising over 24,000 publicly listed North American firms for the period 1962-2014. We find strong evidence for a structural organization and equalization of profit rates on a relatively short time scale both at the economy wide and one- and two-digit SIC industry levels into a Laplace or double exponential distribution. We show that the statistical equilibrium approach is consistent with economic theorizing about profit rates and discuss research questions emerging from this novel look at profit rate distributions. We also highlight the applicability of the underlying principle of maximum entropy for inference in a wide range of economic topics.

Keywords: Firm competition, Laplace distribution, Profit rate, Statistical equilibrium, Principle of Maximum Entropy.
JEL codes: C15, D20, E10, L11

1 Introduction

One of the most pervasive assumptions in contemporary economic theories spanning various traditions is that competition leads to the formation of equal profit rates

*This research benefited from discussions with Duncan Foley as well as from suggestions by Paulo dos Santos, Deepankar Basu, Sid Dalal, Amos Golan, Mishael Milaković, Fernando Rugitsky, Cato Sanford, Markus Schneider and Anwar Shaikh. We are grateful to two anonymous referees for their helpful feedback on an earlier version.

†University of Missouri Kansas City, 203D Manheim, 5120 Rockhill Road, Kansas City, Missouri 64110. Email: scharfenakere@umkc.edu
‡University of Sussex, 346 Jubilee Building, Brighton BN1 9SL, United Kingdom. Email: g.semieniuk@sussex.ac.uk.

for all firms in equilibrium. Yet, it is well understood by most economists that this situation, which predicts a "degenerate" distribution of profit rates, never actually obtains in reality as observed rates of profit tend to perpetually fluctuate. Available data from national accounts, that report only industry or sectoral averages but no higher moments of the distribution, and a theory that stresses the uniformity of profit rates reinforce the focus on a single rate of profit.

The English classical tradition that began with Adam Smith offers an alternative theoretical approach. Instead of neglecting deviations from the mean, the observed distribution of prices and profit rates are seen as inherent to the process of competition and should be considered as the equilibrium state of competition itself. The appropriate method of analysis is, therefore, not that of the average profit rate, but a probabilistic approach that considers the entire spectrum of profit rates at any moment captured by its probability density. From this perspective, the residual indeterminacy of profit rate deviations from the mean is provided with a rich theoretical structure.

A probabilistic approach to analyzing the behavior of economic variables has been proposed under the label of "statistical equilibrium" before to explain the distribution of prices, incomes and other variables. One of the earliest arguments for a probabilistic approach to study profit rates was made by **Farjoun1983** (henceforth FM). Invoking the classical economists, FM argue that the uniformity assumption of profit rates is a "chimera" and "theoretical impossibility." They propose that prices and profit rates should be treated as random variables that are driven not to a deterministic uniformity amongst capitals, but to time-invariant or stationary probability distributions with general forms that are "theoretically ascertainable and empirically verifiable."[1] With respect to the distribution of the rate of profit, the authors go so far to as to claim the "theoretical and empirical evidence suggests that the rate of profit has a so-called gamma distribution."(**Farjoun1983**) However, FM never detail the theoretical rationale for their claims, and they adduce little evidence for the rate of profit having any particular distribution, let alone a gamma one.

In this paper we examine the notion of *statistical equilibrium* in economic theory and show the validity of this type of reasoning applied to the classical political economic theory about non-degenerate profit rate distributions, by studying the empirical distribution of the rate of profit. We explain the precise theoretical footing of FM's approach, which they elide in their book with a vague reference to physics,[2] and show that it is consistent with economic theorizing. Drawing on firm level data for over 24,000 North American firms from 1962 to 2014 we use FM's two conjectures: that profit rates are stationary, and that they are gamma distributed, to discuss the shape of the distribution. We find strong evidence for a stationary profit rate distribution, but show that the gamma hypothesis is unnecessarily restrictive in its assumptions and poorly supported by the data. Rather, we find that the general form of the profit rate distribution approximates the double exponential or Laplace distribution, a distribution commonly found in industrial organization research (**Alfarano2012**; **Bottazzi2003a**; **Bottazzi2003b**; **Bottazzi2006**; **Stanley1996**). A probabilistic approach shows profit rate distributions in a fresh light that leads to asking new questions about competition and the determination of one of the most important variables in economic theory.

The following section addresses the competitive process as a grounds for studying statistical equilibrium, section three outlines the concept of statistical equilibrium and its foundational principle of maximum entropy on which FM's conjectures are based. Section four introduces the dataset with which we explore FM's conjectures and section five shows the results obtained from analysis of the dataset. In section six we offer an alternative more parsimonious distribution based on the principle of maximum entropy, and in section seven we summarize our finding and discuss avenues for further research.

2 Competition as a Disorderly Process

The assumption that all firms receive an equal rate of profit in equilibrium is a common point of departure across much of the spectrum of contemporary economic theory. The concept of equilibrium for profit rates predominantly comes in two forms. In the first one, profit rates are *a priori* uniform amongst perfectly competitive firms in an economy with all resources fully utilized; equilibrium is therefore a tautology. Deviations from uniformity are a result either of "shocks" that create a temporary disequilibrium with convergence back towards uniform rates, or imperfect competition (**Mueller1986**).

The second form can be traced to Alfred Marshall's concept of short- and long-run partial equilibrium (**Marshall1890**). Marshall maintained that while in the short run firms with different cost structures (the payment of profits or "quasi-rents" to capital providers being a cost) may exist, in the long run firms will enter or exit so as to force the price to equal the minimum average factor costs of the incumbent firms, equalizing profit rates as one of the factor costs (**Foley2011**). The divergence of rates of profit from uniformity is treated as a short-run phenomenon that can be abstracted from in the long run (**Kurz1995**). In either case, equilibrium is understood as a state of uniformity.

The theory of equilibrium profit rates balances with the concomitant theory of competition. Indeed, a different theory of competition will lead to an alternative notion of how profit rates will appear over longer periods of time. A case in point is the – as Sraffa pointed out "submerged and forgotten" – approach of the English classical school and their critic, Karl Marx. Adam **Smith1982** began from the notion that it was the competitive disposition of capital to persistently seek higher rates of profit and that through competition a *tendency* to the equalization of profit rates across all competitive industries would emerge. Karl Marx would later summarize competition as a perpetually turbulent and dynamic process whereby "capital withdraws from a sphere with a low rate of profit and wends its way to others that yield higher profit." (**Marx1981**) This "constant migration" of capital leads to a distribution of capital among the various spheres of production and consequently a distribution of profit rates.

From this perspective, the dynamics of profit rate oscillations can be understood as arising from a negative feedback mechanism. Capital seeking out sectors or industries where the profit rate is higher than the economy-wide average, generates new investment in these sectors attracting labor, raising output, and reducing prices and profit rates.[3] On the reverse side, firms' departure from industries earning below an average rate of profit has the opposite effect as the reduction in supply provides an incentive for capital to leave the sector, which leads to higher prices and profit rates for firms that remain in the sector. The important point is that the tendential gravitation of each industry's rate of profit is an unintended consequence of the decisions of many individual firms. That is, firms do not choose their profit rates. It is the entry and exit decisions of other firms in the process of competition that leads to a realization of a profit rate at any point in time. Resolution of the process of competition only emerges at the macroscopic scale as captured by the distribution of the prime mover of capital, the rate of profit.[4] Arguably, Marx was well aware of this microscopic/macroscopic dichotomy as he states,

> "[The] sphere [of circulation] is the sphere of competition, which is subject to accident in each individual case; i.e. where the inner law that prevails through the accidents and governs them is visible only when these accidents are combined in large numbers, so that it remains invisible and incomprehensible to the individual agents of production themselves." (**Marx1981**)

The question then is how to reconcile the ostensible disequilibrium phenomenon of firm level profit rates with a robust notion of equilibrium? FM answer this question by turning to the alternative theoretical framework of classical statistical mechanics which allows them to explain how the observed "chaotic" movement of capital can give rise to a time invariant equilibrium profit rate distribution. Below we outline the intuition behind this notion of statistical equilibrium and its relationship with information theory in order to emphasize this method as one of inference.

3 Statistical Mechanics and Statistical Equilibrium

The methods that Farjoun and Machover promote in *Laws of Chaos* are those of statistical mechanics. While these methods have been well established in physics for more than a century dating back to the foundational work of **Maxwell1860 Boltzmann1871** and **Gibbs1902** they have for the most part been unknown to economists.[5] Important exceptions include the theoretical work of Duncan Foley, who coins the term statistical equilibrium in economics and uses it to model an exchange economy (**Foley1994**) and unemployment in a labor market (**Foley1996**), the "classical econophysics" of **Cottrell2009** who focus on the process of production, exchange, distribution, and finance as a process of interacting physical laws, a number of empirical studies on money, income and wealth distributions such as **Chatterjee2005**; **Dragulescu2000**; **Franke2015**; **Isaac2014**; **Milakovic2003**; **Yakovenko2007**; **Scharfenaker2015**; **Schneider2015**; **Shaikh2014** and the recent work of **Alfarano2012** who apply the principle of maximum entropy to profit rate distributions. Pioneering work in the application of statistical physics to economics is also contained in the work of **Mantegna2000** who explore uses for these methods in analyzing the stock market. The recent work by **Frohlich2013** also follows FM's approach and he finds evidence for gamma distributed profit rates using German input-output tables. All of these works include to some extent the notion of maximum entropy, either in its thermodynamic or its information theoretic form. We outline the basic intuition behind the concept of maximum entropy as a form of inference as in **Jaynes1957a**; **Jaynes1957b**; **Jaynes1979** But in order to get to the inferential approach relevant to the problem at hand, we begin with a basic problem in physics that first led to the use of maximum entropy reasoning.

3.1 Entropy in statistical mechanics

An elementary problem in statistical physics is describing the state of an "ideal" gas, closed off from external influences, that consists of a very large number, N, of identical rapidly moving particles undergoing constant collisions with the walls of their container. The intention is to derive certain thermodynamic features at the macroscopic level (such as temperature or pressure) of the gas from the microscropic features of the gas, that is, from the configuration of all its particles. To describe the microscopic state of this system, however, requires specification of the position and momentum of each particle in 3-dimensional space. These $6N$ coordinates fully describe the microstate the system is in, but for any reasonable system of interest $N \gg 1$ and the degrees of freedom make such a description a formidable task. This problem invites a probabilistic approach to the determination of the microstate.

The state of a particle can be mapped from these $6N$ microscopic degrees of freedom to an energy for the entire system. If we partition the state space of energy levels into discrete "bins" we can categorize individual particles corresponding to their level of energy associated with that particular bin. We can then describe the distribution as a histogram vector $\{n_1, n_2, \ldots, n_k\}$ where k is the number of bins, n_i is the number of particles in bin i, and $\sum_k n_k = N$, the total number of particles (degrees of freedom). The histogram describes the distribution of energy over the N particles but tells us nothing about the exact microscopic state of the system, i.e. the precise location of each particle within each bin. Since there are many *combinations* of particles that lead to the same distribution of particles over bins, any histogram will correspond to many microstates. An *ensemble* is such a partition of the state space with an assignment of probabilities that is a representation of the macroscopic state of the system from which the salient features of the gas are derivable. The macrostate with the largest number of corresponding microstates describes the "statistical equilibrium" of the system. The number of microstates corresponding to any particular macrostate is the "multiplicity" of that macrostate. The insight of statistical mechanics is that macrostates of the gas that can be achieved by a large number of microstates, i.e. have higher multiplicity, are more likely to be observed because they can be realized in a greater number of ways. Because of the combinatorics, the multiplicity of a few macrostates will tend to be much higher than that of all others, effectively allowing for the prediction that the system will be in those highest or near highest multiplicity states most of the time.

To see this, compute the multiplicity of any microstate, that is the number of combinations of particle distributions over bins, which can be expressed as the multinomial coefficient

$$\frac{N!}{n_1!n_2!\ldots n_k!} = \frac{N!}{(N\frac{n_1}{N})!(N\frac{n_2}{N})!\ldots(N\frac{n_k}{N})!} = \frac{N!}{Np_1!Np_2!\ldots Np_k!} \tag{1}$$

The numerator is the number of permutations of N and the denominator factors out all permutations with the same particles in a bin to arrive at combinations only. For large N, the Stirling approximation $\log[N!] \approx N\log N - N$ is a good approximation to the logarithm of the multinomial coefficient.

$$\log[\frac{N!}{n_1!n_2!\ldots n_k!}] \approx N\log[N] - \sum_{i=1}^{k} Np_i\log[Np_i] \tag{2}$$

$$= N\log[N] - N\log[N]\sum_{i=1}^{k} p_i - N\sum_{i=1}^{k} p_i\log[p_i] \tag{3}$$

$$= -N\sum_{i=1}^{k} p_i\log[p_i] \tag{4}$$

This sum is therefore an approximation of the logarithm of the combinations (multiplicities) and is called entropy.[6] The logarithm of the multiplicity is a concave function of the probability $n_k/N = p_i$. Since the probabilities sum to one and since the entropy is a concave function of the multiplicity, the entropy corresponding to the macrostate with the largest multiplicity – entropy at its maximum – occurs when all probabilities are equal, $p_i = k/N,\ \forall i$.

Constraints on the possible configuration of particles can be imposed, since typically there is some knowledge about the system that makes some configurations more or less probable. These tend to take the form of linear functions of the probabilities. For example, imposing the condition that energy is conserved results in the constraint on the mean energy in the system. Maximizing entropy (the concave objective function in Eq. 4), subject to the constraint that the momenta of the particles result in a kinetic energy equal to the total energy of the system coarse-grained into N discrete energy states, $\sum_{j=1}^{N} x_j p_j = \bar{x}$ is a mathematical programming problem that is quite familiar to economists (**Foley2003**).[7]

3.2 The maximum entropy principle of inference

Claude **Shannon1948** working on problems of digital communications at Bell Labs in 1948, had a very practical interest in developing a consistent measure of the "amount of information" in the outcome of a random variable transmitted across a communication channel. He wanted logical and intuitive conditions to be satisfied in constructing a consistent measure of the amount of uncertainty associated with a random variable X using only the probabilities $p_i[x_i],\ x_i \in X$.[8] Shannon discovered that this measure of information, the average length of a message, was identical to the expression for entropy in thermodynamics (Eq. 4). Shannon was interested in assigning probabilities to messages so as to maximize the capacity of a communication channel. The physicist Edwin T. Jaynes (**Jaynes1957a**) recognized the generality of Shannon's result and discovered that this situation is not very different from that in statistical mechanics, where the physicist must assign probabilities to various states. He argued that the number of possibilities in either case was so great that the frequency interpretation of probability would clearly be absurd as a means for assigning probabilities. Instead, he showed that Eq. 4 is a measure of information that is to be understood as a measure of uncertainty or ignorance.[9] That is, the probability assignment in Eq. 4 describes a *state of knowledge* in the sense common to Bayesian reasoning.

From this perspective, Jaynes argued that imposing constraints on systems that change the maximum entropy distribution, such as energy conservation, are just instances of using information for inference. Inferring the maximum entropy distribution subject to information about the system is what Jaynes then referred to as the principle of maximum entropy inference (PME).[10] In an economic context, examples of information we use to impose constraints (closures) are, for example, a budget constraint, the non-negativity of prices, accounting identities, behavioral constraints, or the average profit rate in an economy. So long as these constraints are binding we should expect to find persistent macroeconomic phenomena consistent with the PME. As Jaynes describes the PME, "when we make inferences based on incomplete information, we should draw them from that probability distribution that has the maximum entropy permitted by the information we do have."(**Jaynes1982**)

In applications, inferences about a system from imposed informational constraints are typically made by using moment constraints as with the mean energy constraint. The problem is then one of maximizing entropy subject to the normalization constraint of probabilities summing to one, as well as any other constraints that act on the system as a whole.[11] From this perspective, we can see that for Jaynes the ensemble of "bins" and their relative probabilities describe a certain state of knowledge and for this reason he believed the connection between Shannon's information theory and statistical mechanics was that the former justified the latter.

We stress that the concept of entropy in economics is a general method of inference that is not in direct correspondence with the physical theory of thermodynamics. We use the PME to make inferences about the mechanisms and processes in a complex system with many degree of freedom that generate stable macroscopic regularities. The predictive relevance of maximum entropy inference is conditional on the ability of the statistical model to produce observable regularities in the system under analysis. The statistical model is determined by information about the system incorporated as constraints that modify the predictive distribution. Thus, the probability distribution that achieves maximum entropy over the relevant domain is a logically equivalent representation of our state of knowledge. That is, the maximum entropy distribution expresses what we know while being maximally non-committal toward what we do not.

3.3 Farjoun and Machover's gamma conjecture

FM implicitly follow this logic in arguing that, "[t]he chaotic movement in the market of millions of commodities and of tens of thousands of capitalists chasing each other and competing over finite resources and markets, cannot be captured by the average rate of profit, any more than the global law of the apparently random movement of many millions of molecules in a container of gas can be captured by their average speed." (**Farjoun1983**) They predict a probabilistic approach to better explain this chaotic movement, and derive two claims from it. First, that the distribution should not change its shape or location over time, that is, that the profit rate distribution is stationary. This is the notion of a statistical equilibrium, as opposed to a deterministic one. Secondly, they predict the shape of the distribution by reasoning that "In a gas at equilibrium, the total kinetic energy of all the molecules is a given quantity. It can then be shown that the 'most chaotic' [the maximum entropy] partition of this total kinetic energy among the molecules results in a gamma distribution... if we consider that in any given short period there is a more-or-less fixed amount of social surplus... and that capitalist competition is a very disorderly mechanism for partitioning this surplus among capitalists in the form of profit, then the analogy of statistical mechanics suggests that R [the rate of profit] may also have a gamma distribution."(**Farjoun1983**)

The gamma distribution FM propose is defined for a continuous random quantity $X \in [0, \infty)$ by two parameters $\alpha > 0, \beta > 0$

$$G[x; \alpha, \beta] = \frac{\beta^{\alpha}}{\Gamma[\alpha]} x^{(\alpha-1)} e^{-\beta x} \tag{5}$$

where brackets indicate arguments of functions, α is the shape parameter of the positively skewed distribution, and β is a rate parameter. The expected value of a gamma distribution is $\mathrm{E}[x] = \frac{\alpha}{\beta}$ and the variance $Var[x] = \frac{\alpha}{\beta^2}$. Figure 3 shows the gamma distribution for various β while holding $\alpha = 4$ (**Farjoun1983**).

[Figure 1 about here.]

Using the PME, we can ascertain that FM's claim that the annual profits in an economy should be partitioned so as to give rise to a gamma distribution rests on particular constraints on the entropy maximization program: a non-negative support, and fixed arithmetic and geometric means. In Appendix B, we show that the gamma distribution indeed arises from these three constraints. For these constraints to apply, they should be motivated using economic theory, as they amount to assumption of three instances of *information* about the economy. First, non-negativity of profit rates, secondly, a "conserved" mean profit rate. Thirdly, the geometric mean constraint translates into the constant average *growth rate* of the size of profits for firms of any size or "scale invariance" of profit rates. FM do not spell out nor motivate these assumptions claiming too great "technical and conceptual difficulties" (**Farjoun1983**), but simply claim a gamma distribution in analogy with a gas. Once these assumptions are spelled out however, our theoretical prior is from the beginning that this model is misspecified as the gamma distribution is only defined on the positive support $[0, \infty)$ which *a priori* excludes negative profit rates. We are more agnostic about the other two constraints, which might possibly be justified from economic theory. A look at actual profit rate distributions will test the accuracy of our information based on which we infer a distributional shape.

4 Data and Plot Method

4.1 Data

The data we examine is from the merged Compustat/CRSP Annual Northern American Fundamentals database comprising US stock market-listed companies spanning the years 1962-2014,[12] for which the distribution of profit rate cross sections have not yet been analyzed. Government as well as financial services, real estate, and insurance have been excluded because the former does not partake in competition and the latter adhere to different accounting conventions for revenue calculation that makes this part of the industries incomparable. We calculate the profit rate by dividing the difference of net sales and operating costs, which equals gross profit, by the book value of total assets. From a Marxian point of view, the income accruing to capitalists should ideally be divided by "total capital advanced," which includes fixed assets, raw materials and unused labor power, and commercial and financial capital (**Basu2013**). Total assets is an incomplete, but non-restrictive measure of "capital advanced" that is measured at historical cost. From our theoretical perspective, assets measured at "replacement cost" would be a more consistent measure of capital advanced, however, such a measure is unavailable. Additionally, we do not believe the two measures will deviate in any systematic or meaningful way at this level of analysis.

As for comparability of results with those in FM (chapter 7) this cannot be ascertained since FM do not disclose how their measure of capital is defined. Unlike FM who attempt to partition indivisible units of capital across profit rate bins we choose to partition indivisible firms across profit rate bins due to the structure of the data. We emphasize the fact that Compustat only provides *ex post* year-end annual profit rate observations while the competitive process that generates the equilibrium distribution is *unobserved*. With this in mind, a complete description of the dynamic entry and exit process that gives rise to the equilibrium profit rate distribution is not possible and the problem is, therefore, ill-posed and incomplete. That is, while a firm like Apple Inc. exists in the dataset for 35 years, they have competed in many different niche markets such as phones, computers, televisions, tablets, video game consoles, and watches; where Apple has found themselves in an unprofitable position they have exited the market (e.g. the Apple Pippin). It is the simultaneous decisions of many similar firms to enter and exit particular markets that gives rise to the observed profit rate distribution; however, inference is limited to year end “snapshots” of this complex dynamic process.

Lastly, outliers have been removed using a nonrestrictive Bayesian filter which comprise only 3 per cent of the data. Details about the filter are in (**Semieniuk2015**). Our dataset, therefore, is comprised of firms under the standard industrial classification (SIC) numbers 1000-6000 and 7000-9000, containing a total of 285,698 observations with on average 5,390 annual observations. The summary statistics for the complete data set are presented in Table 1. Appendix A contains full details about constructing the dataset.

[Table 1 about here.]

Firm level profit rate distributions have not yet been analyzed in this manner.[13]

4.2 Logarithmic density plots

When plotting the histograms of profit rates, logarithmic density scales will be used. This is common in analyzing distributional shapes in the industrial organization literature because it facilitates identification of exponential distributions and power laws (**Bottazzi2006**; **Stanley1996**). The effect of rescaling is illustrated in Figure 2. Both plots show the same normal, Laplace, and gamma density where the left plot has a linear density scale and the right one has a logarithmic one. (Double) exponential distributions with their fatter tails than normal ones can be easily identified because they appear as a straight line (where the slope equals the parameter of the exponential distribution), which will be important when analyzing the empirical distributions. The Laplace distribution assumes a tent shape, making it easy to distinguish it from a normal distribution. The gamma distribution's positive tail also approximates a straight line for $\alpha \approx 1$ (in the plot $\alpha = 4$). In log plots of empirical distributions it is important not to exaggerate the significance of variations in the tails, as these are actually very low densities with small variations, which are magnified due to the re-scaling.

[Figure 2 about here.]

5 Empirical Densities

We plot the empirical density of profit rate cross sections for a selection of years in Figure 3 on a log probability scale which emphasizes the tent shape of the distribution. The points correspond to the middle of the histogram bin for a given year. Histograms are stacked in the same plot window with different markers for every cross section plotted at five year intervals, the upper for the years 1965-1980 and the lower one with years spanning 1985-2010.

[Figure 3 about here.]

5.1 Time invariance of the profit rate distribution

What immediately stands out from these figures is first, how remarkably organized the profit rate observations are into a stationary distribution. The empirical densities display a clear tent shape characteristic of the Laplace distribution. We can also see that the majority of firms achieve profit rates between -20 and +50 per cent, but there are important outliers in every year on both sides. In every year the distribution has roughly the same mode, and densities fall off in exactly the same pattern. Slight differences only occur in the tails, but recalling the log scale of the density axis, these variations are very slight indeed and to be expected in noisy, observational data. Clearly, FM's first conjecture – that profit rate distributions are expected to be organized into a time invariant distribution – is well supported by the data. The statistical equilibrium hypothesis is a good one for the time period both in the upper and lower plots.

The only change to the shape of the distribution occurs in the 1980s, when the negative tail swells, resulting from an increasing amount of firms that realize less than the modal profit rate and remain alive and active. This lends a negative skew to the erstwhile symmetric distribution and suggests there are two "eras" spanned by the data. It has been observed that starting in the 1980s, profitability and net worth measures for firms in all industries began to drop (**Peristiani2004**).[14] The profit rate cross sections suggest that the statistical equilibrium is perturbed through a new constraint on firm activity (survival at lower profit rates), but that it reasserts itself in a new form.

Another important observation from Figure 3 is the astounding organization of profit rates above the mode relative to profit rates below the mode. It appears as if the dynamic pressures exerted on firms earning above the most likely, or modal rate of profit leads to far greater stability of the macroscopic state of profit rates. This may suggest that "the general rate of profit," as put forth by classical political economists as a type of "reference rate" with which entry and exit decisions would be determined, might appropriately be captured by the mode of the distribution. This reference rate of profit is typically equated with the average rate of profit and in the case that the distribution is symmetric the mode will coincide with the mean as well as the median. In the later decades of our sample, however, the average lies to the left of the mode due to the asymmetry of the distribution.

In contrast to FM, who use limited sectoral data of the total of British manufacturing industries, the far more comprehensive Compustat dataset shows the distribution of profit rates for almost all industries. Further, the apparent statistical equilibrium Laplace distribution is robust to disaggregation by industry. Plots of industry-wide and sectoral distributions are presented in Appendix C and show that the statistical equilibrium distribution is present at the one- and two-digit SIC sectoral industrial group level.

5.2 The shape of the distribution

The profit rate distribution is distinctly non-normal and hence deviations from the dominant average are not well explained as a sample of independently and identically distributed random variables that tend to a normal distribution. But the distribution is also markedly non-gamma, and this is clear first from the fact that the support of the distribution significantly extends into the negative realm, second, the gamma distribution is positively rather than negatively skewed, and last, the sharp peakedness and tent shape that appears in the log histogram is more characteristic of the Laplace distribution.[15] The empirical evidence appears to contradict FM's conjecture; yet, a careful discussion of their argument yields additional information about the actual shape of the distribution. FM are acutely aware of the non-negativity problem and immediately question "whether it is reasonable to assume that $f_R(r)$ is equal, or very close, to 0 for all negative r. This would mean, in other words, that in a state of equilibrium only a negligible proportion of the total capital has a negative rate of profit (that is, makes a loss)."(**Farjoun1983**) While they claim that "At first sight it would seem that this assumption is quite unrealistic," FM nevertheless argue in favor of the gamma distribution due to their belief that

1. In normal times the proportion of capital (out of the total capital of the economy) in the negative rate-of-profit brackets is much smaller than first impressions suggest.

2. Among the firms that actually do make a loss, there is usually a disproportionately high number of small firms (firms with a small amount of capital). For this reason the proportion of loss-making capital (out of the economy's total capital) is considerably smaller than the number of loss-making firms would suggest.

3. When a firm is reported to be making a 'loss', what is usually meant is a loss after payment of interest on the capital it has borrowed. This is the 'loss' shown in the balance-sheet of the firm. However, for the purpose of comparison with our model, the interest paid by the firm must be taken as part of the profit. A firm whose rate of profit (in our sense) is positive but considerably lower than the current rate of interest, and whose capital is partly borrowed, may end up (after payment of interest) with a net loss on its balance-sheet.[16]

The first point is not well supported in our dataset as firms realizing negative rates of profit comprise an astounding 20 per cent of total observations and in some years over 30 per cent of all profit rates are negative. While there is a clear cyclicality to the percentage of negative profit rate observations to total observations in Figure 4, it appears that the "normal times" of which FM speak may have changed since the time of their writing. The prevalence of negative profit rates may be considered a norm and this trend certainly warrants further research.

[Figure 4 about here.]

The second point, that on average negative profits are made by small firms, is supported by our data until the 1980s, but fails to hold true in the newer decades. We illustrate this in Figure 5 by showing Tukey box plots of profit rates conditional on firm size.[17] We pool all observations for each era and display the distribution of profit rates for each capital percentile.

[Figure 5 about here.]

From Figure 5 we can see that negative profit rates until around 1980 are hardly existent except for very small firms. The lower inter-quartile bound is negative only until the the fifth percentile, meaning more than one quarter of firms in that size bracket realize negative profit rates. This condition changes considerably later on when negative profit rates are pervasive for at least the first four deciles and not until the 45th percentile does the inter-quartile range lie exclusively above zero. Even some of the largest firms suffer from negative profits and the phenomenon of negative profit rates is spread over the entire size distribution in both eras. This mass of firms earning negative profit rates adds the negative asymmetry to the distribution that is evident in Figure 3. The white diamonds are average profit rates conditional on firm size and only coincide with the median and mode above the 5th decile of the size distribution (half of the dataset), which is where the profit rate distribution becomes symmetric. It is important to realize that the shift is primarily in the higher moments of the distribution, and not a shift of the whole distribution and its first moment. However, the skew implies that first moment and the mode of profit rate realizations no longer coincide.

FM's last point that concerns the definition of profit rates is accounted for by our calculation that measures gross profits as total revenue less operating expenses before depreciation. It is from this gross pool of surplus that profit is distributed *post festum* to interest payments on capital, taxes, et cetera, leaving more meagre returns for the production capitalists. But the present analysis shows that some 20 to 35 per cent of firms even struggle with negative *gross* profit rates.

It follows, therefore, that negative profit rates are an important characteristic of the general profit rate distribution and assuming non-negativity would indeed amount to an "extremely rigid" approach that is not, "at all realistic [..] as an approximate description of the behaviour of a real capitalist economy" (**Farjoun1983**). It is also important to realize that non-negativity is a gratuitous assumption in the statistical approach to economic theory. In our discussion of the results in the next section, we show that reasoning based on the PME can equally well lead to a distribution, that can better capture the shape of the statistical equilibrium for negative values. It can be motivated in the same way as we did above for the gamma distribution.

6 An Alternative Theoretical Distribution

Having established that profit rates are not well approximated by a gamma distribution, but pursuing a statistical equilibrium approach, the foregoing visual inspection leads us to propose a Laplace or double exponential distribution as the theoretical distribution that better approximates the shape of the empirical one. The Laplace distribution is the maximum entropy distribution when the mean absolute size of deviations from the average of the quantity under consideration is constrained and the continuous random variable X is non-vanishing on the open support $(-\infty, \infty)$ (**Kotz2001**). From the point of view of economic theory, the constraint leading to the Laplace distribution can be interpreted as competitive pressure on the dispersion of profit rates, without setting an absolute lower boundary such as non-negativity. Firms straying too far above the general rate of profit are reined in by stricter competition from new entrants. Firms sustaining negative profits for too long change sectors or go out of business.

6.1 The Laplace distribution as a candidate

Formally, the Laplace distribution arises from the following maximum entropy programming problem:

$$\begin{aligned} \max_{\{p[x]\geq 0 \,|\, x\in \mathbb{R}\}} \quad & -\int_{-\infty}^{\infty} p[x]\log[p[x]]dx \\ \text{subject to} \quad & \int_{-\infty}^{\infty} |x-\mu|p[x]dx = c \\ \text{and} \quad & \int_{-\infty}^{\infty} p[x]dx = 1 \end{aligned} \tag{6}$$

where $c > 0$ is a constant and $\mu \in \mathbb{R}$ is the location parameter corresponding to a reference rate of profit.[18] When plugging the new constraints into the general form of the mathematical programming problem posed in Appendix B, the solution is the Laplace distribution:

$$L[x; c, \mu] = \frac{1}{2c} e^{-|x-\mu|/c} \tag{7}$$

Visual inspection of the fit suggests that this density function fits the data well for all firms until 1980, as seen in Figure 6, and is surprisingly parsimonious in its single moment constraint. Other studies tend to use distributions with a higher number of parameters, which are theoretically more difficult to motivate (**Buldyrev2007**). Using the PME highlights this issue, as each additional parameter also requires an additional constraint that must be motivated by information about the distribution derived from economic theory. Those constraints that are redundant will not change the maximum entropy distribution and those that are unwarranted or incorrect will lead to poor predictive distributions.

[Figure 6 about here.]

In the newer era we have seen that the negative tail becomes fatter and an asymmetry develops. This translates into a skewed density plot in Figure 7, suggesting that while the symmetric Laplace model does a remarkable job of capturing the major features of the profit rate distribution, there appear to be some secular trends in the higher moments of the distribution that the single constraint Laplace model does not predict.

6.2 Beyond symmetric Laplace

One way of dealing with the asymmetry is to assume that the relevant process generating the equilibrium profit rate distribution is for a subset of larger more established firms only. Smaller firms are then subject to separate constraints. Indeed, the earlier boxplot in Figure 5 showed that above the median firm size, symmetry asserts itself even in the newer period. Pursuing this logic, as did **Alfarano2012** we end up discarding a substantial amount of data and information and can only reason about a minor subset of the population of firms embroiled in competition. This raises the question of how competition among small, and competition among large firms influence each other. On the other hand, assuming that all firms are competing with each other, and following the logic of **Jaynes1979** this innovation in the distribution can be accounted for by additional constraints. We propose one possible avenue for approaching this problem by solving for the maximum entropy distribution with an additional constraint motivated by economic theory, but leave the question open for further research.

[Figure 7 about here.]

In order to incorporate additional information that will modify the maximum entropy distribution we can impose a constraint that will introduce an asymmetry into the distribution. One logical candidate is to treat infra- and supra-modal observations as subject to separate mean constraints. Formally, the constraint becomes a piecewise mean constraint around the modal, reference rate of profit (μ).

$$
\begin{aligned}
\max_{\{p[x]\geq 0 \,|\, x\in \mathbb{R}\}} \quad & -\int_{-\infty}^{\infty} p[x]\log[p[x]]dx \\
\text{subject to} \quad & \int_{-\infty}^{\mu} (\mu - x)p[x]dx = c_1, \\
& \int_{\mu}^{\infty} (x-\mu)p[x]dx = c_2, \\
& \int_{-\infty}^{\infty} p[x]dx = 1
\end{aligned}
\tag{8}
$$

The mathematical solution to this modified programming problem is the asymmetric Laplace distribution (**Kotz2001**; **Kotz2002**):

$$
AL[x;c_1,c_2,\mu] = \frac{1}{(\sqrt{c_1}+\sqrt{c_2})^2}
\begin{cases}
e^{\frac{(x-\mu)}{\sqrt{c_1 c_2}+c_1}} & \text{if } x \leq \mu \\
e^{-\frac{(x-\mu)}{\sqrt{c_1 c_2}+c_2}} & \text{if } x > \mu
\end{cases}
\tag{9}
$$

where $c_1, c_2 > 0$ and $\mu \in \mathbb{R}$. The addition of the piecewise constraint greatly improves the fit of the distribution as is evident in Figure 7. The additional constraint continues to capture the stable exponential character of supra-modal profit rates while also accounting for the asymmetry of the distribution. Although infra-modal profit rates after 1980 no longer appear as a straight line – indicating a trend in the higher moments of the infra-modal part of the distribution – the two constraint maximum entropy asymmetric Laplace specification is remarkably more successful at capturing the general structure of the profit rate distribution throughout the sample.

Formally, these constraints separate positive and negative deviations from the mode into two different components of the aggregate distribution. Theoretically, this parametrization of the asymmetric Laplace distribution is motivated by the idea of separate entry and exit dynamics where firms below the modal, or reference rate of profit, are worked upon by different pressures from the macro environment than those on the positive side. For supra-modal firms, there are strong organizing principles that structure the firm population's ability to generate returns, which result in a statistical equilibrium. That is, entry competition as captured by the c_1 constraint is constantly producing an exponential distribution above the mode or general rate of profit. This results in a stable c_1 over the entire sample period which is seen in Figure 8. On the other hand, the exit pressure on infra-modal firms – captured by the c_2 constraint – appears to be trending downwards, allowing an ever greater deviation in losses for a growing fraction of U.S. stock-market listed firms, that additionally fluctuates at business cycle frequencies.

[Figure 8 about here.]

We believe these constraints are theoretically consistent with a "classical" theory of competition where profitability is defined in reference to a general rate of profit that in turn regulates the distribution of capital. Further, we see no *a priori* reason to believe the pressures acting on capital realizing returns below the general rate to be the same as those acting on capital realizing returns above the general rate. In fact, a reasonable prior would be that firms facing exit pressures, such as the liquidation of used capital stock, face a quite different environment than firms attempting to maintain their relatively profitable position or of new firms entering these relatively profitable markets. This line of reasoning leads to explicit consideration of the *joint* distribution of unobserved entry and exit decisions and the way these decisions indirectly determine the *marginal* distribution of profit rates observed here.[19] When $c_1 = c_2$ the constraint becomes redundant and the maximum entropy distribution collapses into the symmetric Laplace distribution. Interestingly, this appears to be the case prior to 1980, suggesting more "symmetric" pressures across the spectrum of profitability.

Conventional goodness-of-fit tests like the Kolmogorov-Smirnov test show that neither symmetric nor asymmetric Laplace distributions capture all features of the noisy empirical data. We note that these tests do not impose a penalty for additional model complexity and with large samples sizes are highly sensitive to small deviations from the candidate distribution. We favor the application of the more sensitive entropy based *information distinguishability criterion* (**Soofi2002**) that shows that the asymmetric Laplace candidate captures more than 90 per cent of the informational content in the empirical distribution. Explaining the residual indeterminacy would require additional constraints motived by economic theory that avoid overfitting. Details and output of these measures are discussed in Appendix D.

In summary, although we have an incomplete description of the distribution of profit rates in the newer era, we still see a remarkable amount of organization of the distribution in both eras and we believe this organization represents important information about the competitive process, which can largely be captured by a parsimonious maximum entropy model. The apparent change in the equilibrium distribution suggests our knowledge about the newer era may be incomplete and new information needs to be accounted for. The proposed statistical approach allows the researcher to confront and study this phenomenon and challenges economic theory to investigate this development more closely.

6.3 Economics and the Principle of Maximum Entropy

Before concluding, we stress the potential of the PME method of inference for economics. FM claimed loosely that competition between firms could be seen as molecules interacting in a gas, therefore profit rates – like the characteristics of molecules – should be distributed as a gamma distribution. We have shown that starting from maximum entropy as a general principle of inference forces the researcher to be precise about the constraints that lead to a particular maximum entropy distribution and that the predictive relevance of maximum entropy inference is conditional on the ability of the statistical model to produce observable regularities in the system under analysis. The case of the non-negativity assumption for a gamma distribution allowed us to exclude the gamma distribution as a candidate due to strict inability to ever predict a negative profit rate.

Crucially, we believe that in the selection of alternative candidate distributions, additional constraints on the maximum entropy program require economic justification. Insight into competition between firms is not improved simply by observing that the data is better approximated by an asymmetric exponential power distribution than an asymmetric Laplace distribution (Appendix D), which is not surprising given the additional model complexity. The data tell us that there is additional information that is not explained by existing constraints, but theory must be able to motivate constraints that can account for this additional information. This may require developing theoretical insights based on a more granular examination of the irregularities, such as the characteristics of the firms that make the negative tail fatter than the positive one. The joint application of the PME coupled with motivation from economic theory can lead to new research questions and insights about the profit rate. It is obvious that this method of inference may be applied to any quantity in economics for which an ensemble can be constructed.

7 Conclusion

The distribution of the rate of profit is the outcome of an enormous number of independent decisions of individual firms to compete in a multitude of disparate markets. It is an unintended consequence of individual capitals seeking higher rates of return which gives rise to observed statistical regularities. The present research into the empirical distribution of firm profit rates started from the assertion that the general form of the profit rate distribution arising from this disorderly process is neither degenerate nor an arbitrary time variant one. For publicly traded U.S. firms, we have shown that profit rates are extremely well organized into a tent shape characteristic of the Laplace distribution. This distribution shows strong time invariant qualities consistent with the statistical equilibrium hypothesis, displaying only a shift in the higher moments in the 1980s towards a new, skewed invariant distribution.

We have argued for the viability of statistical reasoning and the maximum entropy approach for this problem and have shown that given the right information the configuration of profit rates can be well approximated by a maximum entropy distribution with one to two constraints, both motivated by economic theory. For the dataset that we considered, two paths of further research emerge from the present study that we consider particularly promising for generating new insights. First, identifying the relevant information about competition in the last three and half decades that can account for the reconfiguration of the profit rate distribution and formulating them as proper constraints may shed light on structural changes in the competitive environment of the US economy. This will require more theorizing about why firms that do less well than the general rate of profit can survive longer now than in the period before the 1980s. Secondly, the lack of coincidence of the distribution's first moment and its mode due to the skew of the distribution leads to the question what summary information about profit rates is sufficient to use in analyzing an economy. Just as with heavily skewed income distributions, where the average earning may not say much about the income of most households, a mean profit rate pulled down by negative outliers may hide a "healthier" profit rate of the large majority of firms when doing aggregate industry analysis.

This research, based on applying the principle of maximum entropy and explaining constraints with economic logic, leads to asking new questions about long-existing phenomena, not only for profit rates, but also in a number of other fields. One example is the labor market where information about wages and obstacles to mobility translated into constraints on possible wage distributions may reveal a statistical equilibrium in wage rates dependent on sector or country. Another is the interpretation of firm rates of growth or capital accumulation as a process where information about markets – and firm covariates such as the profit rate – may yield a rich theory of statistical equilibrium in industrial dynamics. The increasing availability of micro-level datasets further encourages analyzing the whole distribution of several economic variables rather than only one or at best two moments, which are not sufficient to capture the information contained in the distribution in case of non-Gaussian distributions.

Notes

[1]See **Jensen1953**; **Pikler1951** for early developments of similar stochastic modeling.

[2]The same can be said of **Frohlich2013** who argues in favor of the gamma distribution, but does not explain the underlying assumptions linking it to a statistical equilibrium framework.

[3]See **Fratini2016** for a fully developed stochastic model of price gravitation from a classical perspective.

[4]See **Shaikh2015** for a discussion of equilibrium as a turbulent equalization process. Shaikh argues that the "law of one price" disequalizes profit rates within an industry due to differing technology, but that capital flows between industries into *regulating capitals* "turbulently equalizes" the profit rate. Equilibrium, in this sense is perpetual fluctuation of the profit rates of regulating capitals (approximated by the "incremental rate of profit") around a common average value.

[5]See **Mirowski1991**

[6]The term was introduced by Rudolph Clausius in the 1850s as a measure of energy dissipation in thermodynamic systems. However, since Claude Shannon's pioneering work on information theory (**Shannon1948**) the term has been used to describe a variety of mathematical expressions and the one above has been referred to as the "classical" Boltzmann-Gibbs-Shannon entropy (**Gorban2010**).

[7]The solution to this problem is the exponential distribution (**Kapur1992**).

[8]These conditions were (1) for a discrete random variable with uniform probabilities the uncertainty should be a monotonically increasing function of the number of outcomes for the random variable, (2) if one splits an outcome category into a hierarchy of functional equations then the uncertainty of the new extended system should be the sum of the uncertainty of the old system plus the uncertainty of the new subsystems weighted by its probability, and (3) the entropy should be a continuous function of the probabilities p_i.

[9]Maximum entropy subject only to the normalization constraint is the uniform distribution. Intuitively this corresponds to maximum uncertainty as each possible state is equally probable. On the other hand, minimum entropy is represented by a degenerate distribution, which intuitively represents maximum certainty, one outcome with a probability of one.

[10]Importantly, this inferential way of approaching entropy requires no additional assumptions about ergodicity, and is thus immune to the criticism of, for instance, Nicholas **GeorgescuRoegen1971** As **Jaynes1979** argued, with "the belief that a probability is not respectable unless it is also a frequency, one attempts a direct calculation of frequencies, or tries to guess the right "statistical assumption" about frequencies, even though the available information does not consist of frequencies, but consists rather of partial knowledge of certain "macroscopic" parameters... and the prediction s desired are not frequencies, but estimates of certain other parameters... The real problem is not to determine frequencies, but to describe one's *state of knowledge* by a probability distribution."

[11]This probability assignment of microstates will then describe the state of knowledge which we have. An important implication of the PME is that in these problems the "imposed macroscopic constraints surely do not determine any unique microscopic state; they ensure only that the state vector is somewhere in the HPM [high probability manifold]... macroscopic experimental conditions still leave billions of microscopic details undetermined" (**Jaynes1979**). That is to say, the aggregate

does not favor one micro state over another, unless information about specific microstates leads to different constraints that might give better macroscopic predictions.

[12]We follow the convention of **Fama1992** who point out that there is a serious selection bias in pre-1962 data that is tilted toward big, historically successful firms.

[13]The only other work to examine firm level profit rate distributions, **Alfarano2012** find near-Laplace profit rate cross sections in a small sample of only long-lived firms from *Thomson Datastream* data. Indeed, sample selection based on a covariate such as age or size for studying the distributions of firm characteristics is justified under the prior belief that small or young firms belong to an entirely different set of data that are subject to separate "entry and exit" dynamics, which do not play a significant role in determining the statistical equilibrium distribution. However, the essentially arbitrary determination of what is "long lived" or "large" may prevent an understanding how the large majority of competitive firm profit rates are distributed, since a large share of firms are small or short-lived. We believe a more flexible method for sample selection that explicitly models the noise and signal can improve this line of research.

[14]It is remarkable that this shift coincides with the transition to what has been called a "neoliberal" economic environment in the US.

[15]The Laplace distribution has also been found to describe firm growth rates (**Bottazzi2003a**; **Bottazzi2003b**; **Bottazzi2006**) and a sample of profit rates in long-lived firms (**Alfarano2012**).

[16]This enumeration cites **Farjoun1983**

[17]Tukey box plots show observations removed more than 1.5 times the length of the interquartile support from either the upper or lower quartile as dots rather than whiskers.

[18]The location parameter μ is a fixed reference rate of profit to which firms compare their individual profit rates. In the theoretical density, μ is equal to the mean, median, and mode due to the symmetry of the Laplace distribution that need not enter the programming problem as an explicit constraint. Any translation of the distribution necessarily does not change its entropy. See section 3.1.2 in **Kapur1989** for a variety of derivations of the maximum entropy Laplace distribution with a location parameter as well as remark 3.4.5 in **Kotz2001**

[19]See **Scharfenaker2015a** for a more thorough discussion of this problem as well as a proposed quantal response model of firm competition.

Appendix A: Data Sources

Data is gathered from the merged Compustat/CRSP Annual Northern American Fundamentals database through the University of Sussex. We extract yearly observations of the variables AT = Total Assets, REVT = total revenue, XOPR = operating cost, SIC = Standard Industry Code, FYR = year, CONM = company name, from 1962 through 2014. Subtracting XOPR from REVT and dividing by AT gives the conventional measure of return on assets (ROA) which we use as proxy for gross profit rates. Total assets are reported according to generally accepted accounting principles (GAAP) and are measured at historical cost.

Our raw data set consists of 467,666 observations of each indicator. Subtracting completely missing values, government and finance, insurance and real estate - government because it is not engaged in a search for profit maximization but pursues other objectives and finance et al. because the different accounting methods where part of income is not recorded in REVT, leading to profit rates almost zero - we impute the remaining missing values under the assumption of values missing at random. Our completed case dataset contains 294,476 observations or roughly 7,177 observations per year. This data contains outliers of profit rates greater than 10^8 and less than -10^5 per cent. Our prior is that these observations are an artifact of the ratio we use to calculate profit rates and are therefore noise. We pass our data set year by year through a Bayesian filter that models the data as a mixture of a "signal" Laplace and a diffuse Gaussian "noise" distribution. This method uses a Gibbs sampler to assign each firm a latent variable with posterior probability distribution of either belonging to the "signal" or "noise." We make an unrestrictive decision by discarding all observation with latent posterior mean below 0.05 per cent chance of belonging to the signal. Using this method we discard only 3 per cent of our data effectively ridding our data set of massive outliers with a minimal loss of information. Due to the likelihood dominance of our data this procedure will have imperceptible effects on the general form of the empirical distribution. The advantage of our Bayesian filter compared to an *ad hoc* truncation of the dataset (e.g. clipping the top and bottom 1 per cent) is that it endogenizes our sample selection based on our prior that large outliers are noise, and explains outliers as generated by a different distribution. Since any data truncation would have to be based on some "trial and error" procedure of adequately ridding "unacceptably" large outliers we prefer an endogenous selection process that makes this prior explicit. The remaining data set contains 285,698 observations. It is important to note that due to the extremely diffuse Gaussian noise component of the mixture filter, filtering only removes observations far out in the tails, and hence does *not* influence the density of the empirical distribution except at the extreme outliers in the top and bottom two percentiles. Full details on the filter can be found in **Semieniuk2015**

Appendix B: The Maximum Entropy Derivation of the Gibbs Distribution

Jaynes' principle of maximum entropy inference can be formalized by first translating the assumed properties of a system into moment constraints and then by maximization of uncertainty subject to these constraints via the entropy functional $H[x] = -\int p[x]\log[p[x]]dx$. A "constraint" is understood as any information that leads one to modify a probability distribution. The problem is then one of constrained optimization.

To attain the general solution to this problem consider a finite set of N polynomials $\{g_i[x]\}$ for a continuous variable $x \in \mathbb{R}$ with an undefined probability density $p[x]$. Define the i^{th}-*moment* M_i as:

$$\int p[x]g_i[x]dx = M_i \qquad i = 1, 2, \ldots, N \tag{2}$$

Setting $g_0[x] \equiv 1$ and $M_0 \equiv 1$ as the constraint corresponding to normalization assures $p[x]$ will be a probability density function. The problem of maximum entropy inference is to find $p[x]$ subject to the requirement that uncertainty ($H[x]$) is maximized subject to information we have expressed as moment constraints. What this achieves is maximal ignorance given the information available and an "insurance policy" against gratuitous assumptions or spurious details unwarranted by the data. The constrained maximization problem is:

$$\begin{aligned} &\max_{\{p[x]\geq 0 \mid x\in X\}} -\int p[x]\log[p[x]]dx \\ &\text{subject to} \quad \int p[x]g_i[x]dx = M_i \end{aligned} \tag{3}$$

Using Lagrange multipliers for each moment constraint form the augmented functional:

$$L \equiv -\int p[x]\log[p[x]]dx - \sum_{i=0}^{N} \lambda_i \left(\int p[x]g_i[x]dx - M_i\right) \tag{4}$$

We find the critical points by differentiating the functional with respect to $p[x]$ according to the Euler-Lagrange equation $\frac{\partial F}{\partial f[x]} - \frac{d}{dx}\left(\frac{\partial F}{\partial f'[x]}\right) = 0$. Noticing the second term is equal to zero we get:

$$\frac{\partial L}{\partial p[x]} = -(1 + \log[p[x]]) - \sum_{i=0}^{N} \lambda_i g_i[x] = 0 \tag{5}$$

Solving for $p[x]$ results in the maximum entropy distribution:

$$p[x] = Z[\lambda]e^{-\sum_{i=1}^{N} \lambda_i g_i[x]} \tag{6}$$

where $Z[\lambda] = e^{1+\lambda_0}$ is the undetermined normalization constant referred to as the partition function. The family of distributions of this form are known as the exponential family and Eq. 6 is called the Gibbs distribution.

Determining the Lagrange multipliers through moment constraints is non-trivial, however, FM's assumption of a maximum entropy gamma distribution is easily attained by maximizing $H[x]$ with a constraint on support $[0, \infty)$, the arithmetic mean ($\bar{x}$), and geometric mean ($\tilde{x}$) of a random variable $X \in \mathbb{R}_{x\geq 0}$.

Appendix C: Empirical Densities by Industry

In this appendix we show the robustness of the profit rate distribution to disaggregation. Each plot in Figure 9 shows the profit rate distribution at the two-digit SIC level embedded in the one-digit SIC industry for each era. At the one-digit level (agriculture, mining, construction, manufacturing, trade and transportation, and services) - some industries are combined for visual clarity - the industry profit rate distribution is plotted as a solid line, while at the two-digit level we maintain the plotting method from above. This way the clustering of industries at the two-digit SIC level around their parent industry is evident.

[Figure 9 about here.]

Appendix D: Distribution Fit Tests

Nonparametric goodness-of-fit tests, such as the Kolmogorov-Smirnov test, reject the hypothesis that annual cross sections of profit rates belong to the class of Laplace distributions for almost every year, and to the Asymmetric Laplace distribution for every year after 1967. In order to get more insight into the comparative goodness of fit of the data to different distributions, we use the more sensitive information theoretic measure of fit called the *information distinguishability statistic* (ID) proposed by **Soofi2002** The ID is a normalized measure of how well a proposed parametric distribution function exploits the informational content of an empirical distribution. The ID is defined as:

$$ID[\boldsymbol{f};\boldsymbol{f}^*|\theta] = 1 - e^{-\Delta H[\boldsymbol{f};\boldsymbol{f}^*|\theta]} \tag{2}$$

where $-\Delta H[\boldsymbol{f};\boldsymbol{f}^*|\theta]$ is the entropy difference between the empirical histogram distribution ($\boldsymbol{f}$) and a parametric functional distribution ($\boldsymbol{f}^*|\theta$) where θ is the maximum likelihood estimates of $\boldsymbol{f}^*$. An $ID[\boldsymbol{f};\boldsymbol{f}^*|\theta] = 0$ implies $\boldsymbol{f}^*$ is a perfect parameterization of $\boldsymbol{f}$, i.e $\boldsymbol{f}^*$ exploits all information contained in the data distribution. If $1 > ID[\boldsymbol{f};\boldsymbol{f}^*|\theta] > 0$ then there remains residual indeterminacy that the model does not explain.

Figure 10 shows the *ID* measure for four distributions and the inset shows the low p-values of the Kolmogorov-Smirnov test for the symmetric and asymmetric Laplace candidates that yields little information about the difference in fit. For the *ID* criterion, on the other hand, we can see that up to 1980 both the Laplace and asymmetric Laplace distribution capture an average of roughly 95 per cent of the informational content of the data. After 1980, the informational content exploited by the Laplace distribution deteriorates sharply. The asymmetric Laplace continues to capture on average roughly 93 per cent of the informational content of the data, while the more complex three parameter asymmetric exponential power distribution on average captures 95 per cent of the information over the entire sample; a minimal informational gain over the more parsimonious model. The additional complexity of the asymmetric exponential power distribution arises through added constraints on the maximum entropy program which we do not believe have a clear theoretical justification.

[Figure 10 about here.]

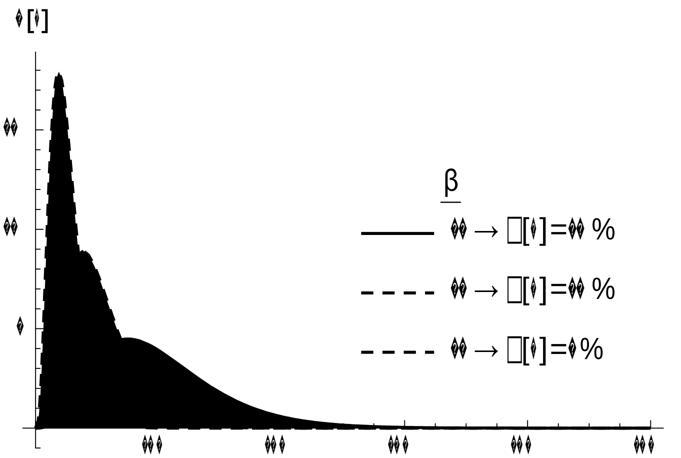

Figure 1: Density plot of the gamma distribution, for $\beta = \{20,\ 40,\ 80\}$ and $\alpha = 4$.

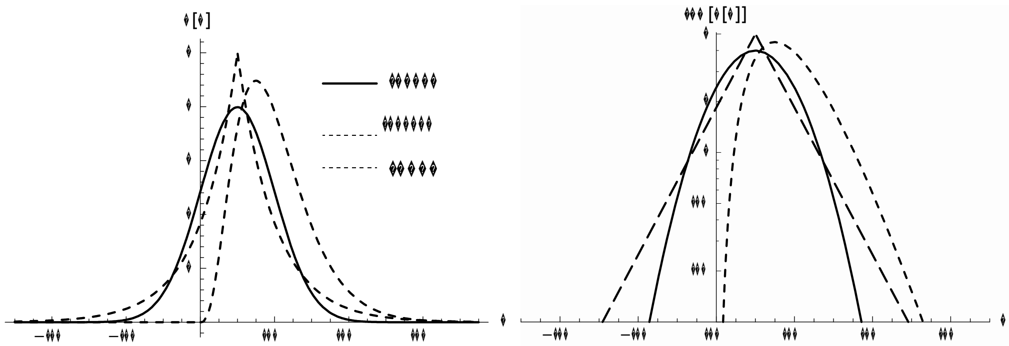

Figure 2: Density plots on a linear (left) and log scale (right) for the normal distribution (solid line), Laplace distribution (dashed line), and gamma distribution (dotted line).

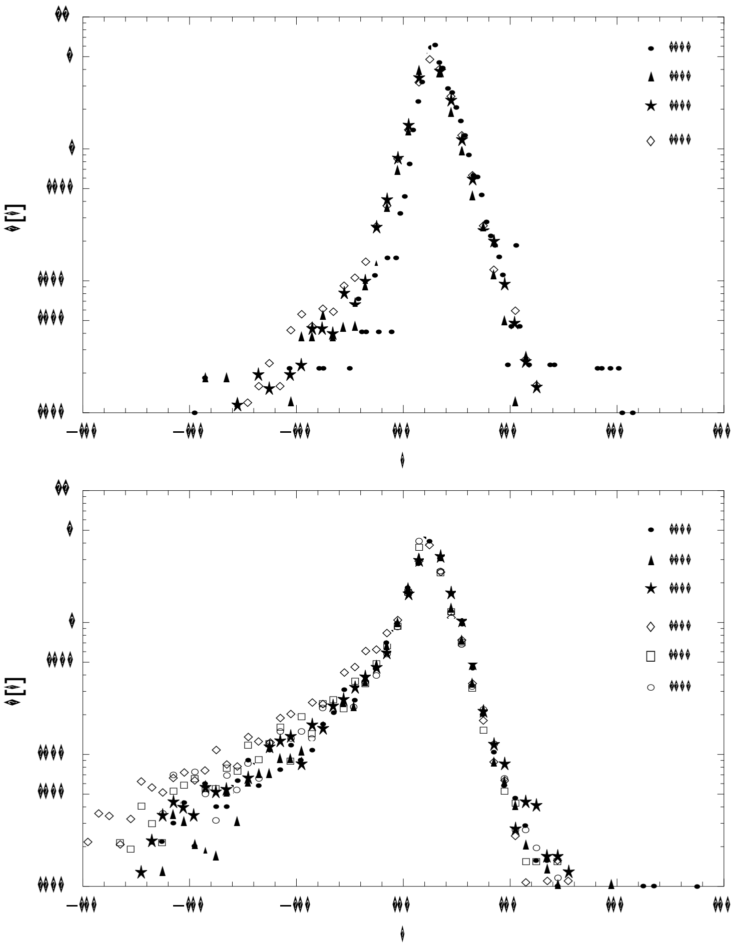

Figure 3: Stacked histogram plots on a log density scale of profit rates (r) for select years. Each shaped point corresponds to the center of the histogram bar for that year. Histograms are stacked in order to show the time invariance of the distribution suggestive of statistical equilibrium.

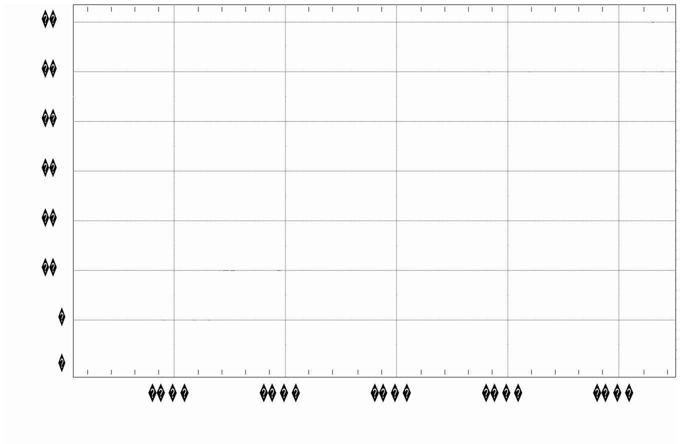

Figure 4: Percentage of negative profit rate observations by year.

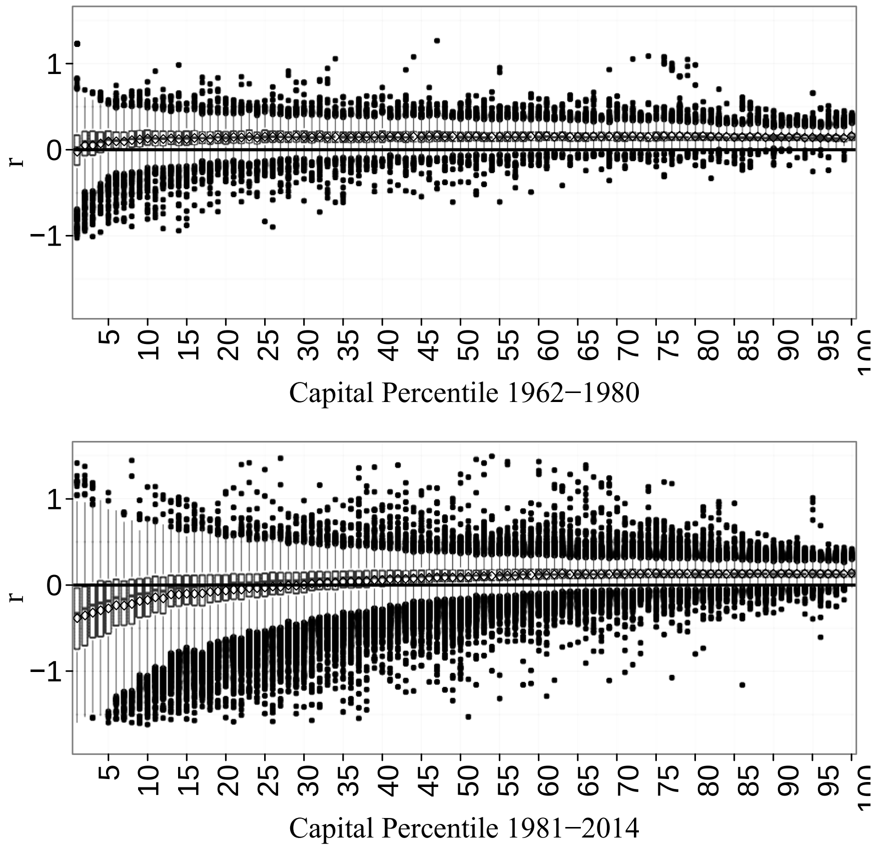

Figure 5: Box plots of profit rates conditional on capital percentile for pooled data between 1962 and 1980 (top) and 1981 to 2014 (bottom). Box plots show the median (black dash), inter-quartile range (box). Outliers appear as points beyond 1.5 times the inter-quartile range. White diamonds are the mean profit rates for that percentile which only correspond to the median for the symmetric distribution.

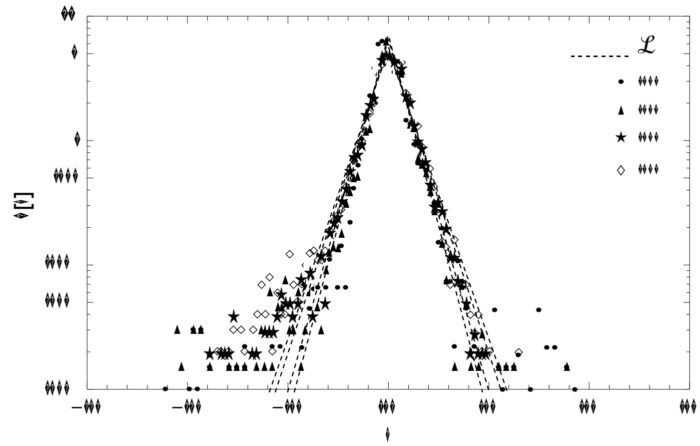

Figure 6: Stacked histogram plots on a log density scale of profit rates (r) for select years for all firms until 1980 with maximum likelihood fitted Laplace distribution (L).

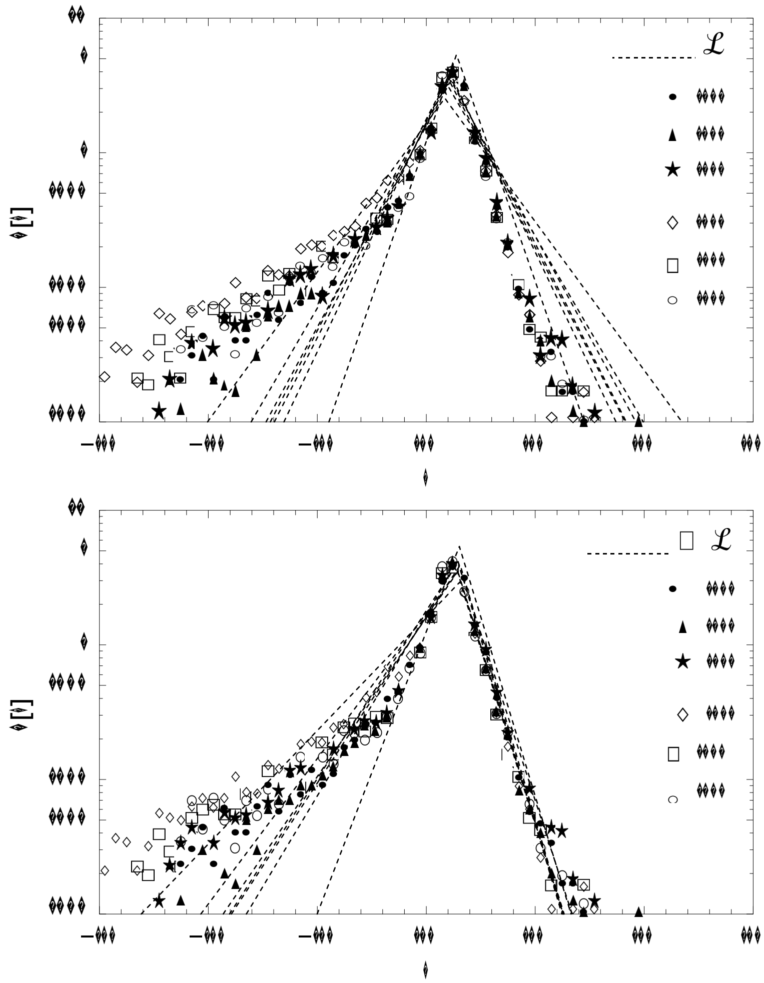

Figure 7: **Top**: Stacked histogram plots for select years after 1980. Dashed lines represent the maximum likelihood fit of the Laplace distribution (L). **Bottom**: Stacked histogram plots for select years after 1980. Dashed lines represent the maximum likelihood fit of the asymmetric Laplace distribution (AL).

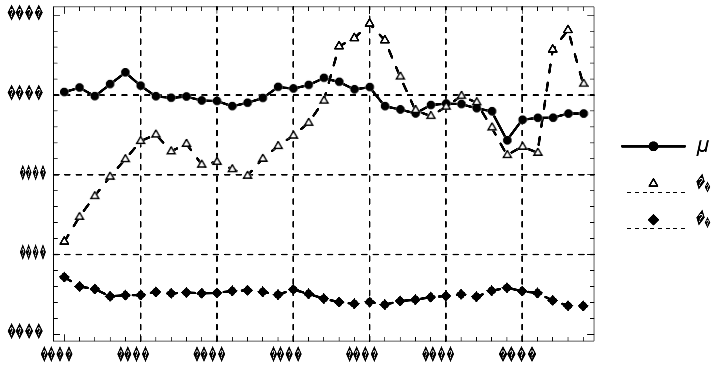

Figure 8: Maximum likelihood estimates of μ, c_1, and c_2 from Eq. 9. Error bars are vanishingly small.

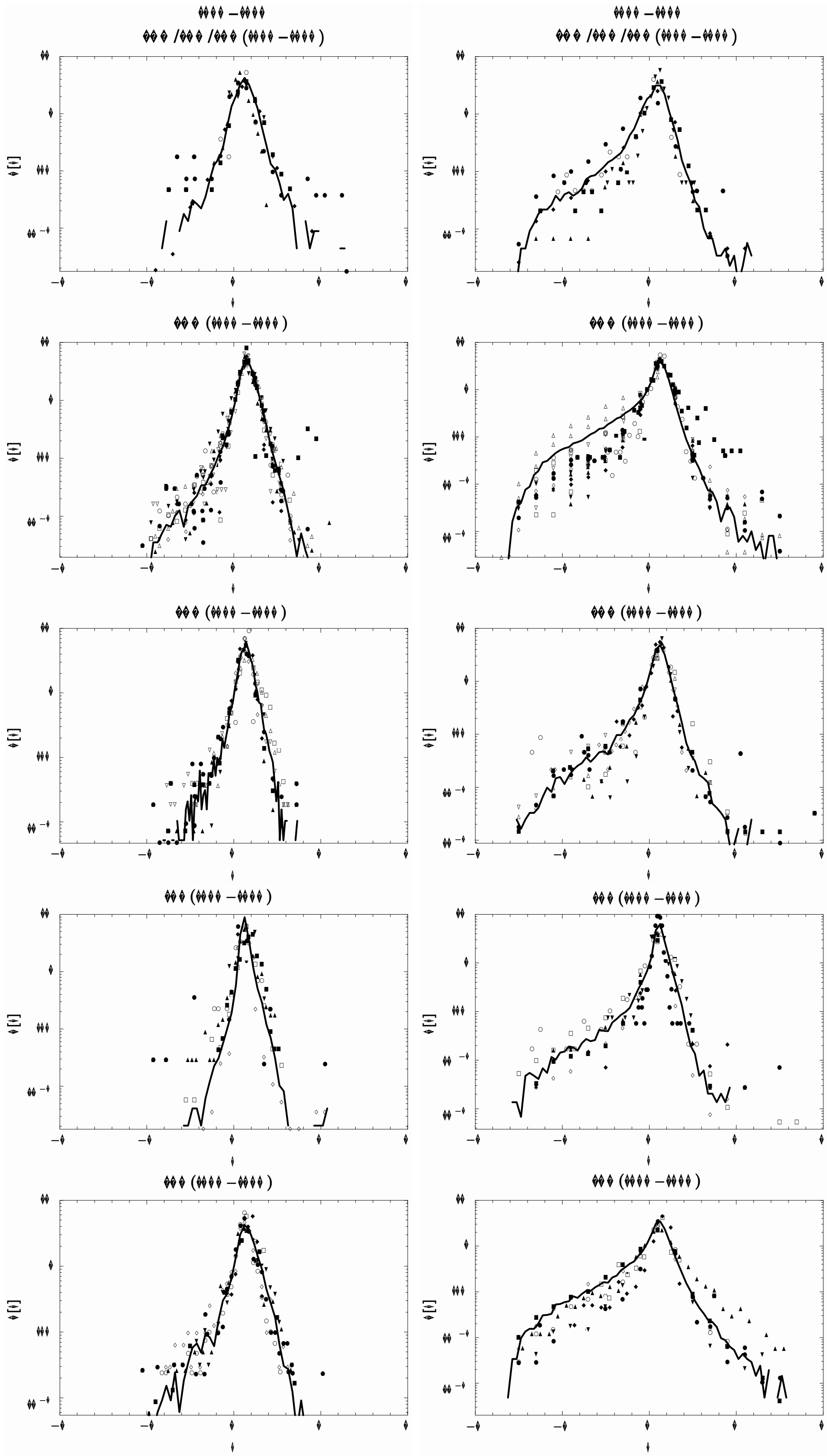

Figure 9: Histogram plots on a log density scale of profit rates (*r*) by 2-digit SIC code. Years 1962-1980 are plotted on the left column and years 1981-2014 are plotted on the right column. Solid lines are the sectoral (1-digit SIC) distribution, e.g. for all of manufacturing firms.

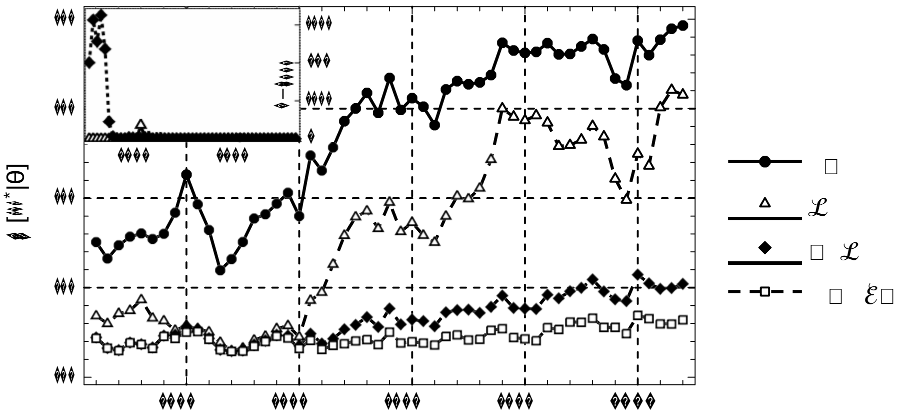

Figure 10: Information distinguishability between four proposed models of increasing complexity. The normal (N) and Laplace (L) are two parameter models, the asymmetric Laplace (AL) is a three parameter model, and the asymmetric exponential power (AEP) is a four parameter model. The inset in the top left shows annual p-values for the Kolmogorov-Smirnov test of goodness of fit of data cross-sections to the Laplace (L) and the asymmetric Laplace (AL) distributions.

Table 1: Summary statistics of pooled profit rates (r).

	Min	5 Perc.	1st Qu.	Median	Mean	3rd Qu.	95 Perc.	Max
r	-1.616	-0.406	0.033	0.112	0.062	0.173	0.294	1.801